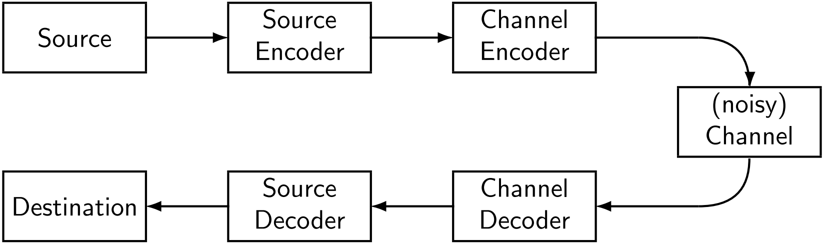
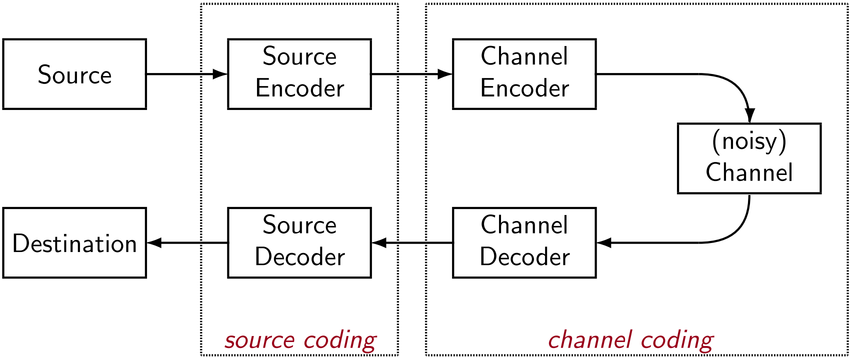
1. Introduction to Channel Coding

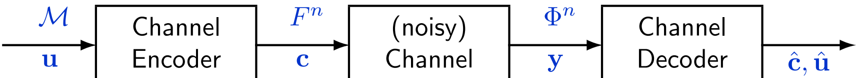
Communication System



Communication System



Channel Coding



Discrete probabilistic channel: (F, Φ, Prob)

- F: finite *input alphabet*, Φ: finite *output alphabet*
- Prob: conditional probability distribution

$$\mathsf{Prob}\{\, \mathbf{y} \text{ received} \mid \mathbf{x} \text{ transmitted} \,\} \quad \mathbf{x} \in F^n, \; \mathbf{y} \in \Phi^n, \; n \geq 1$$

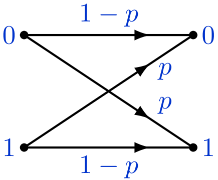
- $\mathbf{u}$: *message word* $\in \mathcal{M}$, set of M possible messages
- $\mathbf{c} \in F^n$: *codeword*
- $\mathcal{E} : \mathbf{u} \xrightarrow{1-1} \mathbf{c}$ *encoding*
- $\mathcal{C} = \{\mathcal{E}(\mathbf{u}) \mid \mathbf{u} \in \mathcal{M}\}$ *code*
- $\mathbf{y} \in \Phi^n$: *received word*
- $\hat{\mathbf{c}}, \hat{\mathbf{u}}$: *decoded codeword, message word*, $\mathbf{y} \longrightarrow \hat{\mathbf{c}} \; (\longrightarrow \hat{\mathbf{u}})$ *decoding*

Code Parameters

$\mathcal{C} = \mathcal{E}(\mathcal{M}) \subseteq F^n, \quad |\mathcal{C}| = M$

- n: *code length*
- $k = \log_{|F|} M = \log_{|F|} |\mathcal{C}|$: *code dimension*
- $R = \frac{k}{n}$: *code rate* ≤ 1
- $r = n - k$: *code redundancy*
- We call $\mathcal{C}$ an (n, M) (*block*) *code* over F

Example: Memoryless Binary Symmetric Channel (BSC)



BSC(p)
$p =$ *crossover probability*

(we can assume $p \leq 1/2$)

- $F = \Phi = \{0, 1\}$
- $\text{Prob}(0|1) = \text{Prob}(1|0) = p, \quad \text{Prob}(0|0) = \text{Prob}(1|1) = 1 - p$
- For $\mathbf{x} \in F^n, \ \mathbf{y} \in \Phi^n$,

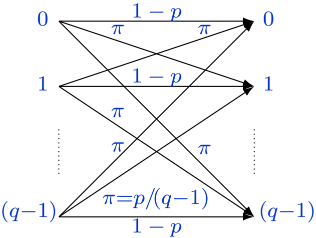
$$\text{Prob}\{\, \mathbf{y} \text{ received} \mid \mathbf{x} \text{ transmitted} \,\} = \prod_{j=1}^{n} \text{Prob}(\ y_j \mid x_j\ ) \ = \ p^t (1-p)^{n-t},$$

where $t = |\{\ j \mid y_j \neq x_j\ \}|$ (number of errors)

Memoryless q-ary Symmetric Channel (QSC)

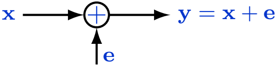
- $F = \Phi, \ \ |F| = q \geq 2$
- For $x, y \in F$,

$$\text{Prob}(y \mid x) = \begin{cases} 1-p, & x = y, \\ \pi \stackrel{\Delta}{=} p/(q-1), & x \neq y. \end{cases}$$



- Assume F is an abelian (commutative) group, e.g.: $\{0, 1, \ldots, q{-}1\}$ with addition mod q.

- *Additive channel* (operating in the group F^n)



- $\mathbf{e} = \mathbf{y} - \mathbf{x}$: *error word statistically independent* of $\mathbf{x}$

$$\mathbf{e} = [\, 0 \ldots 0, e_{i_1}, 0 \ldots 0, e_{i_2}, 0 \ldots 0, e_{i_t}, 0 \ldots 0 \,]$$

$i_1, i_2, \ldots, i_t$: *error locations*	$e_{i_1}, e_{i_2}, \ldots, e_{i_t}$: *error values* $(\neq 0)$

The Hamming Metric

- *Hamming distance*

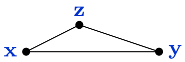
  For single-letters $x, y \in F$: $\mathsf{d}(x,y) = \begin{cases} 0, & x = y, \\ 1, & x \neq y. \end{cases}$

  For vectors $\mathbf{x}, \mathbf{y} \in F^n$: $\mathsf{d}(\mathbf{x},\mathbf{y}) = \sum_{j=0}^{n-1} \mathsf{d}(x_j, y_j)$

  number of locations where the vectors differ

  Example: $\mathbf{x} = (101101001)$, $\mathbf{y} = (110101110)$ $\quad \mathsf{d}(\mathbf{x},\mathbf{y}) = 5\,.$

- The Hamming distance defines a *metric*:
  - $\mathsf{d}(\mathbf{x},\mathbf{y}) \geq 0$, with equality if and only if $\mathbf{x} = \mathbf{y}$
  - Symmetry $\mathsf{d}(\mathbf{x},\mathbf{y}) = \mathsf{d}(\mathbf{y},\mathbf{x})$
  - Triangle inequality: $\mathsf{d}(\mathbf{x},\mathbf{y}) \leq \mathsf{d}(\mathbf{x},\mathbf{z}) + \mathsf{d}(\mathbf{z},\mathbf{y})$



- *Hamming weight* $\mathsf{wt}(\mathbf{e}) = \mathsf{d}(\mathbf{e},\mathbf{0})$ *number of nonzero entries*
- When F is an abelian group, $\mathsf{d}(\mathbf{x},\mathbf{y}) = \mathsf{wt}(\mathbf{x} - \mathbf{y})$

Minimum Distance

- Let $\mathcal{C}$ be an (n, M) code over F, $M > 1$

$$d = \min_{\mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C} \,:\, \mathbf{c}_1 \neq \mathbf{c}_2} \mathsf{d}(\mathbf{c}_1, \mathbf{c}_2)$$

  is called the *minimum distance* of $\mathcal{C}$

- We say that $\mathcal{C}$ is an (n, M, d) code.
- **Example:** $\mathcal{C} = \{000, 111\}$ is the $(3, 2, 3)$ *repetition code* over $F_2=\{0, 1\}$.
  Dimension: $k = \log_2 2 = 1$, rate: $R = k/n = 1/3$.
  In general, $\mathcal{C} = \{00\ldots0,\ 11\ldots1\}$: $(n, 2, n)$ repetition code, $R = 1/n$.
- **Example:** $\mathcal{C} = \{000, 011, 101, 110\}$ is the $(3, 4, 2)$ *parity code* of dimension $k = 2$ and rate $R = 2/3$ over F_2;
  in general, $\mathcal{C}=\{\ (x_0, x_1, \ldots, x_{n-2}, -\sum_{i=0}^{n-2} x_i)\ \}$, $(n, 2^{n-1}, 2)$ over F_2.

Decoding

- $\mathcal{C} : (n, M, d)$ over F, used on channel $S = (F, \Phi, \text{Prob})$
- A *decoder* for $\mathcal{C}$ on S is a function

$$\mathcal{D} : \Phi^n \longrightarrow \mathcal{C}.$$

- *Decoding error probability* of $\mathcal{D}$ is

$$P_{\text{err}} = \max_{\mathbf{c} \in \mathcal{C}} P_{\text{err}}(\mathbf{c}) \ ,$$

where

$$P_{\text{err}}(\mathbf{c}) = \sum_{\mathbf{y}\,:\,\mathcal{D}(\mathbf{y}) \neq \mathbf{c}} \text{Prob}\{\, \mathbf{y} \text{ received} \mid \mathbf{c} \text{ transmitted} \,\} \ .$$

goal: *find encoders (codes) and decoders that make* P_{err} *small*

Decoding example

- **Example:** $\mathcal{C} = \{000, 111\}$, $(3, 2, 3)$ binary repetition code, channel $S = \mathsf{BSC}(p)$. Decoder $\mathcal{D}$ defined by

$$\mathcal{D}(000) = \mathcal{D}(001) = \mathcal{D}(010) = \mathcal{D}(100) = 000$$
$$\mathcal{D}(011) = \mathcal{D}(101) = \mathcal{D}(110) = \mathcal{D}(111) = 111$$

(*majority vote*).

Error probability

$$\begin{aligned} P_{\mathrm{err}} &= P_{\mathrm{err}}(000) = P_{\mathrm{err}}(111) = \binom{3}{2} p^2 (1-p) + \binom{3}{3} p^3 \\ &= 3p^2 - 3p^3 + p^3 \;=\; p - p(1-p)(1-2p)\ . \end{aligned}$$

- $P_{\mathrm{err}} < p$ for $p < 1/2 \Rightarrow$ *coding improved message error probability*
  but information rate is $1/3$!

In general, for the repetition code, we have $P_{\mathrm{err}} \to 0$ *exponentially* (prove!), but $R = 1/n \to 0$ as $n \to \infty$ *—can we do better?*

goal: *find encoders (codes) and decoders that make* P_{err} *small with minimal decrease in information rate*

Maximum Likelihood and Maximum a Posteriori Decoding

- $\mathcal{C} : (n, M, d)$, channel $S : (F, \Phi, \text{Prob})$.
  Maximum likelihood decoder (MLD):

  $$\mathcal{D}_{\text{MLD}}(\mathbf{y}) = \arg\max_{\mathbf{c} \in \mathcal{C}} \text{Prob}\{\, \mathbf{y} \text{ received} \mid \mathbf{c} \text{ transmitted} \,\}, \ \forall \mathbf{y} \in \Phi^n$$

  With a fixed tie resolution policy, $\mathcal{D}_{\text{MLD}}$ is well-defined for $\mathcal{C}$ and S.

- *Maximum a posteriori (MAP) decoder*:

  $$\mathcal{D}_{\text{MAP}}(\mathbf{y}) = \arg\max_{\mathbf{c} \in \mathcal{C}} \text{Prob}\{\, \mathbf{c} \text{ transmitted} \mid \mathbf{y} \text{ received} \,\}, \ \ \forall \mathbf{y} \in \Phi^n$$

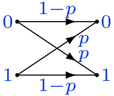
  But,

  $$\begin{aligned} &\text{Prob}\{\, \mathbf{c} \text{ transmitted} \mid \mathbf{y} \text{ received} \,\} \\ &\quad = \ \text{Prob}\{\, \mathbf{y} \text{ received} \mid \mathbf{c} \text{ transmitted} \,\} \cdot \frac{\text{Prob}\{\, \mathbf{c} \text{ transmitted} \,\}}{\text{Prob}\{\, \mathbf{y} \text{ received} \,\}} \end{aligned}$$

  $\Longrightarrow$ MLD and MAP are the same when $\mathbf{c}$ is *uniformly distributed*

MLD on the BSC

- $\mathcal{C} : (n, M, d)$, channel S : BSC(p)



$$\text{Prob}\{\, \mathbf{y} \text{ received} |\, \mathbf{c} \text{ transmitted}\,\} = \prod_{j=1}^{n} \text{Prob}\{\, y_j \text{ received} \mid c_j \text{ transmitted}\,\}$$

$$= \; p^{\mathsf{d}(\mathbf{y},\mathbf{c})}(1-p)^{n-\mathsf{d}(\mathbf{y},\mathbf{c})} = (1-p)^n \cdot \left(\frac{p}{1-p}\right)^{\mathsf{d}(\mathbf{y},\mathbf{c})},$$

where $\mathsf{d}(\mathbf{y}, \mathbf{c})$ is the Hamming distance. Since $p/(1-p) < 1$ for $p < 1/2$, for all $\mathbf{y} \in F_2^n$ we have

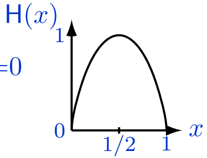
$$\mathcal{D}_{\text{MLD}}(\mathbf{y}) = \arg\min_{\mathbf{c} \in \mathcal{C}} \mathsf{d}(\mathbf{y}, \mathbf{c})$$

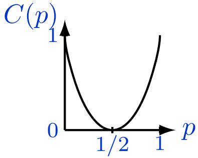
$\mathcal{D}_{\text{MLD}} =$ *nearest-codeword decoder*

- True also for QSC(p) whenever $p < 1 - 1/q$

Capacity of the BSC

- *Binary entropy function* $\mathsf{H} : [0,1] \to [0,1]$
  $\mathsf{H}(x) = -x \log_2 x - (1-x)\log_2(1-x)$, $\mathsf{H}(0)=\mathsf{H}(1)=0$





- *Capacity* of $\mathsf{BSC}(p)$ is given by $C(p) = 1 - \mathsf{H}(p)$

- A special case of the *capacity of a probabilistic channel*, as defined by Shannon (1948)

Shannon Coding Theorems for the BSC

Theorem (Shannon Coding Theorem for the BSC—1948)

Let $S = \mathit{BSC}(p)$ and let R be a real number in the range $0 \leq R < C(p)$. There exists an infinite sequence of (n_i, M_i) block codes over F_2, $i = 1, 2, \cdots$, such that $(\log_2 M_i)/n_i \geq R$ and, for MLD for those codes (with respect to S), the probability $P_{\text{err}} \to 0$ as $i \to \infty$.

Proof. By a *random coding* argument. *Non-constructive!*

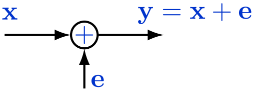
Theorem (Shannon Converse Coding Theorem for the BSC—1948)

Let $S = \mathit{BSC}(p)$ and let $R > C(p)$. Consider any infinite sequence $\{\mathcal{C}_i : (n_i, M_i)\}$ of block codes over F_2, $i = 1, 2, \cdots$, such that $(\log_2 M_i)/n_i \geq R$ and $n_1 < n_2 < \cdots < n_i < \cdots$. Then, for any decoding scheme for $\{\mathcal{C}_i\}$ (with respect to S), the probability $P_{\text{err}} \to 1$ as $i \to \infty$.

Proof. (Loose argument.)

Error Correction

$\mathbf{e} = [0 \ldots 0, e_{i_1}, 0 \ldots 0, e_{i_2}, 0 \ldots 0, e_{i_t}, 0 \ldots 0]$



$i_1, i_2, \ldots, i_t$: *error locations*	$e_{i_1}, e_{i_2}, \ldots, e_{i_t}$: *error values* $(\neq 0)$

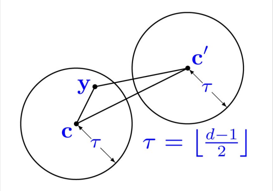
Full error correction: the task of recovering all $\{i_j\}$ and $\{e_{i_j}\}$ given $\mathbf{y}$

Theorem

Let $\mathcal{C}$ be an (n, M, d) code over F. There is a decoder $\mathcal{D} : F^n \to \mathcal{C}$ that recovers correctly every pattern of up to $\lfloor (d-1)/2 \rfloor$ errors for every channel $S = (F, F, \text{Prob})$.

Proof. Let $\mathcal{D}$ be a nearest-codeword decoder. Use triangle inequality. □

Theorem is tight: *For every $\mathcal{D}$ there is a codeword $\mathbf{c} \in \mathcal{C}$ and $\mathbf{y} \in F^n$ such that $\mathbf{d}(\mathbf{y}, \mathbf{c}) \leq \lfloor (d+1)/2 \rfloor$ and $\mathcal{D}(\mathbf{y}) \neq \mathbf{c}$.*



Error Correction Examples

- Binary $(n, 2, n)$ repetition code. Nearest-codeword decoding corrects up to $\lfloor (n-1)/2 \rfloor$ errors (take majority vote).
- Binary $(n, 2^{n-1}, 2)$ parity code cannot correct single errors: $(11100 \ldots 0)$ is at distance 1 from codewords $(11000 \ldots 0)$ and $(10100 \ldots 0)$

Error Detection

- Generalize the definition of a decoder to $\mathcal{D} : F^n \to \mathcal{C} \cup \{\text{`E'}\}$, where $\text{`E'}$ means *"I found errors, but don't know what they are"*

Theorem

Let $\mathcal{C}$ be an (n, M, d) code over F. There is a decoder $\mathcal{D} : F^n \to \mathcal{C} \cup \{\text{`E'}\}$ that detects (correctly) every pattern of up to $d-1$ errors.

Proof. $\mathcal{D}(\mathbf{y}) = \begin{cases} \mathbf{y} & \text{if } \mathbf{y} \in \mathcal{C} \\ \text{`E'} & \text{otherwise} \end{cases}$.

Example: Binary $(n, 2^{n-1}, 2)$ parity code can detect single errors (a single bit error maps an even parity word to an odd parity one)

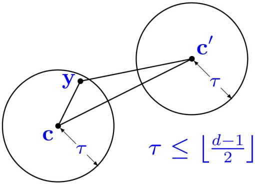
Combined correction/detection

Theorem

Let τ *and* σ *be nonnegative integers such that* $2\tau + \sigma \leq d - 1$. *There is a decoder* $\mathcal{D} : F^n \to \mathcal{C} \cup \{\text{'E'}\}$ *such that*

- *if the number of errors is* τ *or less, then the errors will be recovered correctly;*
- *otherwise, if the number of errors is* $\tau + \sigma$ *or less, then they will be detected.*

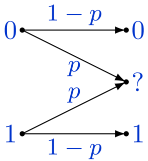
Proof. $$\mathcal{D}(\mathbf{y}) = \begin{cases} \mathbf{c} & \text{if there is } \mathbf{c} \in \mathcal{C} \text{ such that } \mathrm{d}(\mathbf{y}, \mathbf{c}) \leq \tau \\ \text{'E'} & \text{otherwise} \end{cases} .$$



$\tau \leq \left\lfloor \frac{d-1}{2} \right\rfloor$

Erasure Correction

- *Erasure*: an error of which we know the *location* but not the *value*
  $[\, y_1 \ldots y_{i_1-1}, \boxed{?}, y_{i_1+1} \ldots y_{i_2-1}, \boxed{?}, y_{i_2+1} \ldots, \boxed{?}, y_{i_t+1} \ldots y_n \,]$
- *Erasure channel*: $S = (F, \Phi, \mathsf{Prob})$ with $\Phi = F \cup \{?\}$.



Theorem

Let $\mathcal{C}$ be an (n, M, d) code over F and let $\Phi = F \cup \{?\}$. There is a decoder $\mathcal{D} : \Phi^n \to \mathcal{C} \cup \{\text{'E'}\}$ that recovers every pattern of up to $d-1$ erasures.

Proof. On $\rho \leq d-1$ erasures, try all $|F|^\rho$ vectors that coincide with $\mathbf{y}$ in non-erased locations. Find unique codeword, if any. Otherwise, fail (return 'E').

Combined correction/erasure/detection

Theorem

Let $\mathcal{C}$ be an (n, M, d) code over F and let $S = (F, \Phi, \text{Prob})$ be a channel with $\Phi = F \cup \{?\}$. For each number ρ of erasures in the range $0 \leq \rho \leq d{-}1$, let $\tau = \tau_\rho$ and $\sigma = \sigma_\rho$ be nonnegative integers such that $2\tau + \sigma + \rho \leq d{-}1$. There is a $\mathcal{D} : \Phi^n \to \mathcal{C} \cup \{\text{'E'}\}$ such that

- *if the number of errors (excluding erasures) is τ or less, then all the errors and erasures will be recovered correctly;*
- *otherwise, if the number of errors is $\tau + \sigma$ or less, then the decoder will return 'E'.*

- Full error correction "costs" twice as much as detection or erasure correction. Price list:
  - full error to correct: requires 2 units of distance
  - erasure to correct: requires 1 unit of distance
  - full error to detect: requires 1 unit of distance
- How does distance "cost" translate to redundancy "cost"?

Summary

- (n, M, d) *code* over alphabet F:

$$\mathcal{C} \subseteq F^n, \quad |\mathcal{C}| = M, \quad d = \min_{\mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C}, \mathbf{c}_1 \neq \mathbf{c}_2} \mathsf{d}(\mathbf{c}_1, \mathbf{c}_2)$$

- n: *code length*

  $k = \log_{|F|} M$: *code dimension*

  $r = n - k$: *code redundancy*

  $R = k/n$: *code rate*

- *Maximum likelihood decoding*:

$$\hat{\mathbf{c}} = \arg\max_{\mathbf{c} \in \mathcal{C}} \mathsf{Prob}\{\mathbf{y} \ \ \mathsf{received} \mid \mathbf{c} \ \ \mathsf{sent}\ \}$$

- For QSC, equivalent to $\hat{c} = \arg\min_{\mathbf{c} \in \mathcal{C}} \mathsf{d}(\mathbf{y}, \mathbf{c})$ *nearest codeword decoding*

Summary

- Shannon: there are sequences of codes $\mathcal{C}_i(n_i, M_i)$ that allow $P_{\mathrm{err}}(\mathcal{C}_i) \overset{i\to\infty}{\to} 0$ while keeping $R_i \geq R > 0$, as long as $R < C$, where C is a number that depends solely on the channel (*channel capacity*)
  Error-free communication is possible at positive information rates
  (he just didn't tell us how to implement this in practice)
- Maximum likelihood decoding may be too complex: sometimes we need to settle for less
- If $2\tau + \rho + \sigma \leq d - 1$, an (n, M, d) code can
  - correct ρ *erasures* and τ *full errors*
  - *detect* between $\tau + 1$ and $\tau + \sigma$ errors (in addition to ρ erasures)
- Challenges: how to find good codes (codes with large d), how to represent them compactly, how to encode, how to decode